DOCUMENT RESUME

ED 320 913 TM 014 380

AUTHOR McManus, Walter
TITLE Evaluating Adult Literacy and Language Training Programs.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE 31 Oct 83
NOTE 27p.; Prepared for the Literacy and Language Program.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adult Education; *Adult Literacy; Cost Effectiveness; *Data Collection; Education Work Relationship; Employment Programs; English (Second Language); *English Instruction; Equations (Mathematics); *Evaluation Methods; Evaluation Problems; Income; Labor Force Development; Mathematical Models; *Program Evaluation; *Research Methodology
IDENTIFIERS Comprehensive Employment and Training Act

ABSTRACT

Modeling and data collection efforts needed to develop a test of the effectiveness of adult literacy and language training programs in increasing the earnings of trainees are reviewed. After reviewing the framework developed for the Manpower Development and Training Act and the Comprehensive Employment and Training Act (CETA) evaluations and the data sets used to test the models, literacy and language programs are considered in an attempt to determine the information needed for evaluation. Problems addressed include: (1) the post hoc nature of evaluations, precluding the possibility of a true experimental design; (2) the need for large sample sizes to detect small anticipated program effects on earnings of program participants; (3) the necessity for longitudinal data on earnings; and (4) the need to contain the costs of an evaluation. An interesting recent approach is the framework using recursive earnings equations developed by O. Ashenfelter (1975, 1978, 1979). A model, developed by N. Kiefer (1978) to extend Ashenfelter's framework, allows for the possibility that language training might alter the functional form relating human capital and earnings. The Continuous Longitudinal Manpower Survey, a study of demographic characteristics of CETA participants and program impacts on employment and earnings, is reviewed. The use of Current Population Survey data in combination with Social Security Administration data is also covered. Recommendations are provided for the selection of control groups. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320913

Evaluating Adult Literacy and Language Training Programs

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Walter McManus
Department of Economics
University of Florida, Gainesville

October 31, 1983

Prepared for

Literacy and Language Program
Stephen Reder, Director

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204
(503) 248-6800

Preparation of this paper was supported by a contract to the Northwest Regional Educational Laboratory from the National Institute of Education. The opinions expressed in this report do not necessarily reflect the position of the Institute, and no official endorsement by the Institute should be inferred.

TM014380

BEST COPY AVAILABLE

Introduction

A major purpose of adult education and training programs is to improve the well-being of those participating. Participants can range from illiterate refugees trying to learn new skills which will facilitate their adjustment and possibly net them their first job in this country, to undereducated, underemployed native-born Americans trying to upgrade their skills with the hope of getting a better job. Although the costs of such programs are generally known and the outcomes (such as increased employment, decreased dependence on public assistance, better trained employees, etc.) can be measured, until recently the effectiveness of adult education and training programs has not been carefully analyzed in terms of costs and benefits.

Government-sponsored English literacy and language training programs, for example, are introduced as a means of helping member of minority language groups develop English literacy and language skills that will presumably speed their economic assimilation. These new skills are intended to enhance employment opportunities for new immigrants (or other persons with scant English), resulting in a reduction of poverty and a reduction in income inequality. Though such training programs have been around for some time, very little is known about the actual effects of training on program participants. In view of the (perceived) growing importance of non-English speaking immigrants and of the fact that alternative methods of speeding economic assimilation exist, it seems reasonable to ask how effective English literacy and language training programs are in increasing the earnings of trainees.

In recent years, since the late '70's, a group of economist has

begun to address the issue of evaluating the effectiveness of government-sponsored training programs under the Manpower Development and Training Act (MDTA) and the Comprehensive Employment and Training Act (CETA). The evaluation techniques that have been developed can, with appropriate modification, be applied to literacy and language training programs. The purpose of this paper is to suggest further modelling and/or data collection effort needed to develop a test of the effectiveness of adult literacy and language training programs. I start by reviewing the framework developed for MDTA and CETA evaluation and the data sets which have been used to test the models. Then, I turn to literacy and language programs and attempt answers to the questions: What information is needed to determine a program's effectiveness? How easy is it to obtain this information from existing sources? What options exist for new surveys?

Review of the Literature

Knowledge of appropriate methodologies for program evaluation has developed rapidly in recent years. Particularly important work was done by Ashenfelter (1975, 1978, 1979). Ashenfelter faces at least four fundamental problems that confront anyone attempting to estimate the impact of training programs on earnings. First, evaluations are usually carried out "after the fact" and the researcher does not have the luxury of utilizing a true experimental design with random assignment of individuals to "treatment" and "control" groups. Instead the researcher must construct the comparison group and devise a methodology to control for the differences between the groups, since the comparison group is usually not equivalent to the trainee group.

Second, large sample sizes are required to detect the small anticipated program effects on earnings, due to the high variance of earnings. Third, evaluation studies must use longitudinal data on individuals' earnings over a number of year so the researcher can estimate how rapidly the program effects appreciate or depreciate with time. Finally, evaluation studies should seek to obtain data at the lowest cost.

Ashenfelter attempts to address these problems by using the Social Security Administration's (SSA) Continuous Work History Sample (CWHS) as a comparison group for individuals who received MDTA training in 1964. The Social Security Administration maintains a summary year-by-year earnings history for each worker with a Social Security account over the period since 1950. This data may be used, under the appropriate confidentiality restrictions, for research purposes. The CWHS is a random sample of longitudinal earnings

records maintained by the SSA for general research purposes.

Ashenfelter assumes that earnings of program participants and of CWHS comparison group members are generated by the same earnings function. His basic analytic framework involves estimation of recursive earnings equations of the form,

$$(1) \qquad Y_{t+k,i} = \sum_{j=1}^{m} B_j Y_{t-j,i} + B_{m+1} d_i + \epsilon_{it}.$$

Where Y_t is reported SSA earnings in year t, d_i is a dichotomous indicator of trainee status which is 1 for trainees and zero otherwise; k indicates the number of year after training the earnings figure refers to, the B's represent parameters to be estimated. The inclusion of lagged values of earnings on the right-hand side of equation (1) is an attempt to control for differences in individuals' expected earnings in the absence of the program and, if (1) is estimated by ordinary least squares, then $\hat{B}_{m+1}$ is an estimate of how much the average trainee's earnings have been raised by the program.

Ashenfelter provides a discussion of possible ways to justify equation (1) on theoretical grounds. He argues that earnings equations of the form of equation (1) can be generated from models of optimal accumulation of human capital in which prior levels of earnings are proxies for an individual's stock of human capital in the absence of participation in the training program. In the data Ashenfelter uses, the typical trainee shows a drop in earnings in the year prior to training relative to the comparison group. There are two extreme ways to interpret this relative decline in earnings: (1) as a purely transitory decline which is no relationship to the amount

of human capital stock the individual enters training with or (2) the relative decline may be permanent and represent a real reduction in the trainee's human capital stock. In the first case earnings in the year prior to training should be omitted from the regression and in the second they should be included. These two possibilities bound the true effect of training: omitting immediate prior year earnings will tend to underestimate the effect of training while including it will overestimate the effect.

Another problem introduced by Ashenfelter's specification and the pre-program earnings decline is the introduction of a (negative) correlation between an individual's earnings before the program (Y_{t-1}) and participation in the program (d_i). Also note that in equation (1) the residuals (ε_{it}) are typically correlated over time for a given individual. Thus, the program variable (d_i) will be correlated with the residual (ε_{it}) and hence ordinary least squares will lead to biased estimates of the program's impact. This is the problem of selection bias. Fortunately, econometric techniques which take account of both the autocorrelation in the residual and the fact that program participation is endogenous are now available (see Heckman 1978).

Ashenfelter restricts the earnings model to be linear in equation (1). Theoretical consideration do not suggest any specific functional form for earning equations. This defect can be remedied by considering the sensitivity of the results to functional form. In addition, Ashenfelter assumes that participants and non-participants face the same functional form before and after training, and only differ in their intercepts in equation (1). From my own work with

language effects on earnings (McManus, et. al. 1983) I conclude that language skills actually interact with other human capital elements multiplicatively rather then simply additively. This means that, at least in applying Ashenfelter-type models to language training programs, the researcher ought to allow the possibility that training might alter the functional form relating human capital and earnings.

Several important advances in the technology of evaluation proposed by Ashenfelter were made by Kiefer (1978, 1978a, 1979b, 1979c). Kiefer (1978) analyzes a data set specifically designed for this purpose. The data are from a two-and-a-half year longitudinal study of four federally-sponsored training programs undertaken in 1969 for the Office of Economics Opportunity (OEO) and the Department of Labor (DOL). Trainees in ten major SMSAs were sampled. The comparison group was constructed by matching members of the sample of the trainee group, on age, race, and sex, to individuals from the pool of people in the SMSA who were eligible but did not participate. Participants in programs and matched comparison group members were interviewed in four waves; the first as soon as possible after entering the program, the second when the trainee left the program, the third four months later, and the last eight month later. Retrospective data on employment earnings, and hours of work were collected along with demographic information. Thus there was no need to match SSA earnings information for the data. Note, however, that at most one year is allowed for an effect in this data set, whereas SSA data potentially allows a longer time frame for older program cohorts. Of course, trainee cohort could be followed for any number of years, but that would increase costs.

In his analysis of the OEO/DOL data, Kiefer (1978) take an approach which realizes and exploits the fact that trainees and controls are not identical, that assignment to training may not be random, and that some individuals are not employed (have zero earnings). He breaks the effect of training down into an effect on potential earnings and an effect on employment. Thus, Kiefer goes a long way toward resolving some of the problems left by Ashenfelter.

Kiefer (1978) applies a model only to men. The effect of training is measured as a function of the number of weeks of training received. This is a generalization of the procedure of using a dummy dichotomous variable to simply indicate program participation. Post training earnings, Y_i, is given by expected earnings without training plus a quadratic in the number of weeks spent in training. Expected earnings without training is specified as a linear function of pre-training expected earnings, Y_{iL}. So Kiefer uses the following earnings generation function:

$$(2) \qquad Y_i = \alpha_0 + \alpha_1 Y_{iL} + \alpha_2 \text{ WEEKS} + \alpha_3 \text{ WEEKS}^2 + U_{1i}.$$

The variable Y_{iL} is constructed using pre-training earnings and employment for trainees and non-trainees. It represents the earnings an individual could have expected based on his personal characteristics without correction for the probability of unemployment. Equation (2) is estimated only over employed men. If earnings and employment probability are correlated the expectation of U_1 conditional on employment is non-zero; specifically, $E(U_{1i} \mid \text{employment of } i^{th} \text{ person}) = \rho(\sigma^*)^{\frac{1}{2}} \lambda_i$ where ρ is the correlation between the error in (2) and the error in the employment relation, σ^* is the variance of U_1, and $\lambda_i = f(\phi_i)/(1 - F(\phi_i))$ where f

is the standard normal density, F is the standard normal cumulative density and ϕ_i is the predicted value from a probit of employment on independent variables. That is, λ_i are variables that can be constructed from the employment relation, and $[\rho(\sigma^*)^{\frac{1}{2}}]$ becomes a parameter to be estimated. The probit employment relation used by Kiefer is

$$(3) \qquad S_i = \gamma_0 + \gamma_1 Y_{iL} + \gamma_2 \text{WEEKS}_i + \gamma_3 \text{WEEKS}_i^2 + \gamma_4 \text{MAR}.$$

The variable S is not observed but its sign is known. MAR is marital status. The equations (2) and (3) shows what Kiefer means when he says he breaks the effect of training into an employment and a potential earnings effect. Weeks of training appear in both relationships.

Kiefer (1978) also makes the number of weeks a person is trained an endogenous variable. He writes

$$(4) \qquad \text{WEEKS}_i = \beta_0 + \beta_1 Y_{iL} + \beta_2 \text{MAR} + \beta_3 \text{CHILDREN} + \beta_4 \text{AGE} + \beta_5 \text{AGE}^2,$$

where β_0 captures benefits of training considerations common across all individuals, lagged earnings reflect time opportunity costs, marital status and number of children are included to control for their effects on costs, and the age relationship adjusts for differing lengths of benefit period. A natural way to estimate (4) is with all persons, trainees and controls, using Tobit estimation techniques. The predicted value of WEEKS can then be used in equations (2) and (3).

Kiefer (1978) constructs the series on lagged earnings, Y_{iL}, from observations on earnings and employment three quarters before training. He is assuming that earnings one quarter before training

contain a significant transitory (negative) component, but that earnings three quarters before training do not. Predicted values from the following equation are used as the lagged earnings variable in previous equations:

$$(5) \qquad Y_{iL} = \delta_0 + \delta_1 AGE_i + \delta_2 AGE_i^2 + \delta_3 EDUC_i + \delta_4 EDUC_i\ AGE_i + \delta_5 MAR_i + \delta_6 CHILDREN.$$

the dependent variable is observed earnings and the sample consists of males employed three quarters prior to training. A correction is made for the non-zero expectation of the error term in equation (5) and the basis of a pretraining employment probit.

Kiefer (1979a) follows the above framework except for two differences: a dichotomous indicator for program participation is used rather than WEEKS, and the sample is restricted to women.

Keifer (1979b) looks at the men in the OEO/DOL data again from a different perspective. Kiefer considers a model of earnings over time with individual and time-specific effects without assuming these effects are orthogonal to participation in MDTA training. The possibility that selection into a training program may be associated with an unobserved "fixed effect" which also affects the wage, that is ability, is investigated. The association could go either way - low-ability people might select themselves into a program or a program manager might select high-ability potential trainees. The employment-potential earnings made in Kiefer (1978) is dropped, but it should be kept in the back of our minds as we read.

Kiefer (1979b) starts with a basic specification of a model of wage-rate determination over time of the form

$$(6) \qquad Y_{it} = X_i\beta_t + d_{it}\alpha_t + \varepsilon_{it}$$

where Y_{it} is the wage of the i^{th} individual in the t^{th} period, X_i is a vector of personal characteristics that do not change over the period of analysis, d_{it} is the trainee status indicator (zero for all individuals in the pretraining periods and one for trainees in posttraining periods), β_t and α_t are parameter vectors that could vary with time, and ε_{it} is the error term. Kiefer simplifies things by assuming that β and α are constant over time and that $\varepsilon_{it} = \gamma_t + U_{it}$ with the errors (U_{it}) orthogonel to the regressors and uncorrelated across individuals after including time effects (γ_t).

Individual effects are represented by decomposing the error term: $U_{it} = f_i + V_{it}$. If we are willing to assume f_i uncorrelated with other regressors then this becomes a special case of equation (6). Suppose, however, that the fixed individual effect, f_i, is not in fact uncorrelated with d_{it} and X_i. The "ability bias" that results can be circumvented. The model becomes

$$(7) \qquad Y_{it} = X_i \beta + d_{it} \alpha + \gamma_t + f_i + V_{it}$$

where β and α have been assumed constant across time periods, γ_t is the time effect in the t^{th} period, f_i is the individual effect of the i^{th} individual, and V_{it} is the error term. Taking means over time and rewriting (7) in terms of deviations from means gives

$$(8) \qquad Y_{it} - Y_i = (d_{it} - d_i.)\, \alpha + (\gamma_t - \gamma.) + V_{it} - V_i.$$

where the dot subscript denotes means over time and within an individual. Equation (8) does not involve the fixed effects and can be estimated by generalized least squares, constrained so that the coefficient α is the same in each equation (there are T-1 equations for each individual). Intuitively, data on each individual is used to

fix the level of his wage profile, and the trainee-control contrast is used to estimate the effect of training.

Kiefer (1979b) also considers the possibility that training is endogenous. He uses a simple instrumental variables approach rather than the more sophisticated training participation probit of his earlier work.

Kiefer (1979c) extends the analysis Kiefer does for classroom MDTA training to three other government-sponsored training program: (1) the Job Opportunities in the Business Sector (JOBS) program, (2) the Job Corps (JC), and (3) the Neighborhood Youth Corps (NYC). The JOBS program involved on-the-job training. The Job Corps provided basic education as well as skill training and work experience in residental centers. The NYC provided remedial education, skill training, jobs and supportive services for low-income youth; little actual training occurred.

Kiefer (1979c) is an attempt to study these four programs using common assumptions, methods, and data. Previous studies of these programs were not comparable because studies of each of the programs used different assumptions, different comparison groups, and different methods of analysis with various sets of data collected by different means.

The data Kiefer (1979c) uses were collected under contract from OEO/DOL. Trainees were sampled from ten SMCAs, with matched non-trainees also being sampled. As with the MDTA data Kiefer analyzed in previous studies, the trainees and matched controls were interviewed in four waves. The first was as soon as possible after entering the program; the second, when the trainee left the program;

the third, four months after the second; and the last, eight months after the third. In addition, the data were augumented by the SSA records of earnings since 1951 through 1974 (many of the trainees had no earnings prior to the training). Matching of trainees and non-trainees was on the basis of age, race, and sex. With data from the interviews, Kiefer is able to do a study similar to his earlier studies, and with the SSA data he can do a study similar to Ashenfelter.

In a sense, Kiefer (1979c) represents a step backward in evaluation technology. He does not consider the issue of program selection, though he does break effects of training into earnings (potential) and employment components.

Goodfellow (1979) analyzes the same data as Kiefer (1979c), and does not add much to the techniques of analysis. He basically only considers a different functional form for the effects of training than Kiefer.

The programs which originated under MDTA in 1962 were reformulated in 1973 by the Comprehensive Employment and Training Act (CETA), and further modified by amendments to CETA in 1978. Beginning in 1975 the Employment and Training Administration contracted with Westat, Inc. to collect data for a sample of CETA participants (and non-participants). The survey Westat, Inc. developed is called Continuous Longitudinal Manpower Survey (CLMS). Studies using the CLMS have not been published in the economics literature as yet, but the survey design is of interest.

The CLMS (as describes in Westat, Inc. (1977)) is a continuing longitudinal study begun in January 1975 of demographic

characteristics of CETA participants and of program impacts on employment and earnings. Data are collected from prime program sponsor records and from interviews conducted each quarter with new enrollees in CETA programs and from follow up interviews at various intervals up to 36 months after enrollment. A comparison group (separate from the CLMS sample) was obtained from the Bureau of the Census Current Population Survey (CPS). Social Security Administration information is used to compile earnings data for both the comparison CPS group and the CETA enrollees sampled for CLMS.

One study using the CLMS/CPS/SSA data was preformed by the Congressional Budget Office and the National Commission for Employment Policy (CBL/NCEP (1982)). Their analysis of earnings effects of CETA was an extension of the fixed effect model of heterogenous earnings functions used by Kiefer (1979b). The fixed-effect model specifies one person-specific parameter for each individual to account for unique characteristics that cannot be measured directly. The CBO/NCEP model, however, specifies two person-specific parameters per individual to account for unmeasured factors affecting both the underlying level and the change over time in individual long-run earnings potential.

The effect of CETA training was estimated in the CBO/NCEP study from the following:

$$(9) \qquad Y_{it} = \alpha_i + \beta_i t + \gamma \cdot T_{it} + \sum_j \delta_j X_{ji} + \varepsilon_t + \varepsilon_{it}$$

and

$$(10) \qquad \varepsilon_{it} = \rho \varepsilon_{i,t-1} + V_{it}$$

where Y_{it} is earnings in year t of person i, t is time, T_{it} is a

program participation indicator, X are other personal characteristics, ε_t is a year-specific error component, ε_{it} is a person-specific and year-specific error component. The α's, β's, δ's, γ, and ρ are parameters to be estimated. γ is the effect of CETA training on earnings. Equation (9) specifies separate earnings trends for each person in the sample. The individual specific error component is serially correlated with parameter ρ.

The CBO/NCEP (1982) study also ignored the question of selection into the CETA program, as Kiefer (1979b) did. They also ignored employment effects except for women participants. The authors of CBO/NCEP argued that their model was superior to the autoregressive earnings function popularized by Ashenfelter (1975, 1978, 1979) because autoregressive models do not fully compensate for differences in the average pre-program earnings of different groups.

The use of CPS matched controls for the CLMS gives rise to a statistical problem referred to as sample contamination. There is some chance that CPS individuals are CETA participants, and they are not identified. However, the problem is probably negligible since the percentage of such unidentified participants is extremely small. This problem is even more negligible with small language training programs.

There have been many evaluations of manpower training programs, and this survey is not intended to be exhaustive. What I have attempted to do is give a general idea of the kinds of questions and data that have been used in the studies of manpower training programs. These can serve as guides to evaluation of literacy and language training programs.

Before turning to the other parts of this paper there is one

other issue that needs to be mentioned: the question of sample size. How large a sample of participants (and of non-participants) does a researcher need for an adequate analysis? This issue was not really addressed by the empirical studies of manpower training programs discussed here, largely because the researchers were confronted simply with a completed sample and asked to work from there.

Stafford (1979) investigates the question of for what sample sizes, if any, marginal program evaluation costs (including the collection of post-program data for trainees and controls as well as costs of analysis) are less than or equal to the marginal benefits the policy decision-maker gains from the evaluation information. In the context of training programs, the evaluation information is the effect of the program on earnings and is more useful, the more precisely it is estimated. Stafford considers the optimal simultaneous choice in a Bayesian statistical framework of the variance of the estimated program earnings effect, the overall sample size, and the participant-control distribution of the sample.

Pitcher (1979) examines optimal sample-size determination in a classical statistical framework. Conlisk (1979) highlights the correspondence between Stafford's Bayesian and Pitcher's classical approach and discusses the estimation of program effects given a large cross section of short time series data.

Potential Data Sources

I have discussed data used in manpower program evaluation above fairly extensively. I have reviewed the existing available data sources. The use of SSA data to augument data already collected has certain confidentiality requirements, but has been allowed on earlier

projects. The CPS extraction of a control group is potentially useful for creating controls for already existing trainee groups. With SSA and CPS together it might even be possible to conduct an acceptable evaluation of a program that is past history. A control group could be created from the CPS and then earnings data from SSA could be added to both trainee and CPS records. I should point out, however, that I am not sure what this procedure would cost.

Further Modelling and/or Data Collection Suggested

To a large extent the degree of modelling complexity depends on the questions the researchers wishes answers to and on the ability of available data to support complex models. In addition the type of data necessary depends on the questions asked and on the models to be estimated. In general, evaluations of literacy and language training programs ought to follow fairly closely the framework developed by analysts of other manpower training programs. This means that attention ought to center on earnings and employment effects of training, with issues of selection into the program, preprogram earnings downturn, lagged structure of earnings functions, and so on being considered where necessary to obtain precise and believable estimates of program effects.

To satisfy these modelling requirements we generally need to consider three or four main types of relationships: (1) earnings functions, (2) employment relationships, (3) program continuation relations, and (4) program selection. To estimate earnings effects of programs the research needs data on earnings before and after training for trainees. Equally importantly,the researcher also needs data on earnings over the same time period for a non-trainee control group.

Information on before and after employment is needed to estimate employment effects, and thus to break total earnings effects into effects on potential earnings and on employment probabilities. Not all who begin literacy or language training complete the program. They may choose to leave if they gain so much early on that there is no need to stay. That means that early drop-outs may be either the highest or the lowest achievers. To determine which is the case is an empirical problem that requires data on earnings and employment of participants and drop outs. Note that because of the ambiguity concerning the relative achievement of drop-outs that they do not appear to represent a very reliable control group.

Questions of program selectivity are also empirical questions and can be addressed with data on the demographic characteristics of participants and non-participants prior to training.

Modelling of the effects of literacy and language training on earnings can follow the framework laid out by Ashenfelter and Kiefer. We now have econometric techniques to simultaneously deal with estimating the earning effects, employment effects, continuation probabilities, and program selection. Specification search for appropriate relationships can be guided by reflection as well as the results of earlier manpower research. A large part of the modelling effort will depend on what data are used.

Data Collection

For an adequate data set to evaluate the effectiveness of literacy and language training on earnings and employment we need repeated observations of earnings and employment status before and after training for trainees and for the same time period for a control

group. There are basically two ways to construct such a data set: (1) by repeatedly interviewing participants and controls, and (2) by augumenting demographic data on participants (presumably already collected in interviews) with SSA earnings data and constructing a control group from interviews of some other source and obtaining their SSA data also. Both methods have good and bad features.

Repeatedly interviewing participants and controls is a costly process. In addition the accuracy of retrospective data from very long ago is doubtful, so that interviews probably only give accurate information on earnings for a few years. Another drawback of the interview option is that earlier cohorts of training program participants are lost to analysis. If they were not asked retrospective earnings questions when they started training it would be very hard to get that information today.

On the positive side, interview questions can be structured to get at economically meaningful concepts of earnings and employment. That is, the researcher has control over the definitions of terms used in the analysis. The researcher can also obtain detailed information about literacy and language skills and backgrounds of the trainees and controls, which might be useful in the analysis of program effects.

The SSA data also have drawbacks. They require certain confidentiality restrictions be maintained, which may necessitate sending researchers to Washington, D.C. to collect the data. The earnings data maintained by the SSA refer only to employment in jobs covered by Social Security laws, and cannot exceed the legal maximum covered earnings. In addition, both the extent of SSA coverage and the legal maximum reported earnings have not remained fixed over time.

The SSA data do not address the question of whether or not a person is employed, so some other way of getting at employment effects would need to be developed (perhaps assume a person is employed if he has SSA earnings greater than, say, $50 per quarter).

On the positive side the SSA provide a fairly objective measure of earnings comparable across individual. Data are available for the period since 1951, so the researcher would not need to rely on an individual's recollection of his earnings in a distant past year. All of the changes in coverage and maximum reported earnings have been documented and the data can be adjusted to account for these changes. It may be possible to get around the ceiling in reported earnings by using quarterly data: We could estimate annual earnings for a person who attained the SSA maximum earnings in the second quarter by multiplying his first quarter earnings by four, for example.

Any study that attempts to evaluate the effectiveness of training programs needs to obtain at least some data by actually interviewing the trainees. This interview should obtain demographic data: race, sex, age, literacy experience, schooling attainment, ethnic and language background, and national and local origin, at a minimum. Even if the SSA option is selected, it would also be useful to obtain, from the interview, retrospective earnings data. These could then be compared with the SSA earnings data as a validity check.

Control Group Selection

The selection of an appropriate control group is a much more difficult problem than estimating program effects. In a sense, the selection of a control group determines the effects that will be

estimated. This is because the control group is used in the prediction of what participants would have experienced in their careers in the absence of training. For simplicity, suppose that we are interested in estimating linear earnings effects. That is, earnings are characterized by the addition of the program's effects and the contribution of everything else to earnings. We can write this as

(11) $Y = \alpha T + F(X) + u$

where Y is earnings; T is an indicator of training; f(X) is a function that measures the contribution of X to earnings; X represents the other factors than training that we expect to influence Y; and u is an error term.

The idea employed by Westat Inc. is simple: find a sample of non-participants (from the CPS) who are so similar to participants in X that the average values of f(X) will be close for the two groups. However, simple ideas might not yield the best results. Whether trying to match participants and non-participants on the basis of their observed personal characteristics, X, depends on whether the function $f(\cdot)$ can be accurately specified.

If the function $f(\cdot)$ can be thought of as well-specified, which simply means that we are certain that the functional form of $f(\cdot)$ applies equally to trainees and non-trainees for all ranges of X, then direct estimation using all the data, not simply a matched sample, is preferred. Using all the data (by which we mean all the CPS control group data) results in estimators with lower variance than matched-sample estimators yield.

Matching reduces estimation efficiency when the specification is known, and may not be justified when the model is not well-specified

either. Omitted variables are the most commonly cited reason for concern that the model is mis-specified and results in potential estimation bias. Suppose that in addition to the characteristics which we do observe X, we know that another unobserved variable, Z, also affects income. If Z in uncorrelated with T, the treatment variable, after controlling for X, then there is no bias in the estimate of the effect of T on income. In this case there is no gain from matching, and matching results in an efficiency loss. Matching is thought to be justified on the basis of the claim that persons matched on observables are likely to be similar in terms of unobservables as well. While this may be true it is irrelevant; regression on the full sample has the same ability to control for both measured and unmeasured variables as matching. The full sample regressions will continue to yield more precise estimates of program impact. If, on the other hand, Z is correlated with T, then estimates of program effects will be biased. Unfortunately, matching does not remove the bias.

Matching and full-sample regression are both extreme solutions. The matching method treats only the matched control group information as informative while the full-sample regression method treats matched and unmatched control group information as equally informative. Both can be viewed as special cases of weighted regression. In the full-sample method equal weights are applied to all control observations. In the matching method equal positive weights are applied to matched observations and zero weights are applied to all other control observations. Between these extremes lie a range of estimation techniques that attempt to use all the data available but

do not assume the same earnings model fits training participants and the whole population.

References

Ashenfelter, Orley, "The Effect of Manpower Training on Earnings: Preliminary Results," in Proceedings of the Twenty-Seventh Annual Winter Meeting of the Industrial Relations Research Association (Madison: IRRA, 1975).

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings," Review of Economics and Statistics, Feb. 1978.

Ashenfelter, Orley, "Estimating the Effect of Training Programs on Earnings With Longitudinal Data," Research in Labor Economics: Supplement 1, Evaluating Manpower Training Programs, 1979.

Congressional Budget Office and National Commission for Employment Policy, CETA Training Programs: Do They Work for Adults? (Washington, D.C., 1982).

Conlisk, John, "Choice of Sample Size in Evaluating Manpower Programs: Comments on Pitcher and Stafford," Research in Labor Economics, Supplement 1, 1979.

Goodfellow, Gordon, "Estimates of the Benefits of Training for Four Manpower Training Programs," Research in Labor Economics, Supplement 1, 1979.

Heckman, James, "Dummy Endogenous Variables in Simultaneous Equation System," Econometrica, July, 1978.

Kiefer, Nicholas, "Federally Subsidized Occupational Training and the Employment and Earnings of Male Trainees," Journal of Econometrics 8 (1978).

Kiefer, Nicholas, "Training Programs and the Employment and Earnings of Black Women," in Lloyd, Andrews, and Gilroy Eds., Women in the Labor Market (New York: Columbia University Press, 1979) a.

Kiefer, Nicholas, "Population Heterogeneity and Inference from Panel Data on the Effects of Vacational Education," Journal of Political Economy, 87:5 p. 2. 1979, b.

Kiefer, Nicholas, "The Economic Benefits from Four Government Training Programs," Research in Labor Economics, Supplement 1, 1979, c.

McManus, Walter, William Gould, and Finis Welch, "Earnings of Hispanic Men: The Role of English Language Proficiency," Journal of Labor Economics, April, 1983.

Pitcher, Hugh, "A Sensitivity Analysis to Determine Sample Sizes for Performing Impact Evaluation of the CETA Programs," Research in Labor Economics, Supplement 1, 1979.

Stafford, Frank, "A Decision Theoretic Approach to th Evaluation of Training Programs," Research in Labor Economics, Supplement 1, 1979.

Westat, Inc., Continuous Longitudinal Manpower Survey: Technical Report No. 1, Methodology, 1977.